DOCUMENT RESUME

ED 469 044 EC 309 184

AUTHOR	Stinson, Michael; Elliot, Lisa; McKee, Barbara; Coyne, Gina
TITLE	Speech Recognition as a Support Service for Deaf and Hard of Hearing Students: Adaptation and Evaluation. Final Report to Spencer Foundation.
INSTITUTION	Rochester Inst. of Technology, NY. National Technical Inst. for the Deaf.
SPONS AGENCY	Spencer Foundation, Chicago, IL.
PUB DATE	2002-06-17
NOTE	18p.
CONTRACT	199800208
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Assistive Technology; *Classroom Communication; College Students; *Computer Assisted Instruction; *Computer Mediated Communication; Computer Uses in Education; *Deafness; *Hearing Impairments; Higher Education; Participant Satisfaction; Student Attitudes; Teaching Methods
IDENTIFIERS	*Automatic Speech Recognition

ABSTRACT

This report discusses a project that adapted new automatic speech recognition (ASR) technology to provide real-time speech-to-text transcription as a support service for students who are deaf and hard of hearing (D/HH). In this system, as the teacher speaks, a hearing intermediary, or captionist, dictates into the speech recognition system in a computer that converts the dictated words of the teacher into print. The process of the captionist repeating the teacher's words is called shadowing. One purpose of the project was to adapt this process to ASR so that it could function successfully as a support service. The second purpose was to evaluate the effectiveness of the ASR system. This evaluation focused on the ability of 10 students who are D/HH to remember information presented with ASR and the perceptions of the students regarding the extent they could comprehend information that was presented with the ASR system. Results from the evaluation tentatively suggest that students did better on recall and recognition tests with an interpreted presentation than with the ASR presentation. Students also indicated that the lag-time, the need to wait for sentences to be displayed as text, interfered with, that is "bothered," comprehension. (CR)

Reproductions supplied by EDRS are the best that can be made
from the original document.

EC

Speech Recognition as a Support Service for Deaf and Hard of Hearing Students: Adaptation and Evaluation
Grant 199800208

Final Report to Spencer Foundation
June 17, 2002

Michael Stinson, Lisa Elliot, Barbara McKee, and Gina Coyne

National Technical Institute for the Deaf
Rochester Institute of Technology

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
R. Davila
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 469 044

Purpose of Project

This project adapted new automatic speech recognition (ASR) technology to provide real-time speech-to-text transcription as a support service for deaf and hard of hearing (D/HH) students. In this system, as the teacher speaks, a hearing intermediary, or captionist, dictates into the speech recognition system in a computer that converts the dictated words of the teacher into print. The process of the captionist repeating the teacher's words is called "shadowing." One of the purposes of the project was to conduct this work to adapt ASR so that it could function successfully as a support service.

The second purpose of the project was to evaluate the effectiveness of the ASR system. This evaluation focused on the ability of D/HH students to remember information presented with ASR and the perceptions of the students regarding the extent they could comprehend information that was presented with the ASR system.

This report has three parts. The first part focuses on the work to apply ASR to the support of D/HH students in regular classes. The second part discusses the evaluation. The third part discusses and interprets the outcomes of these two components of the work and makes suggestions for further work.

Adaptation of ASR

Overview

Work in adapting ASR began with simply learning to dictate from reading printed material into the ASR system with high accuracy. Subsequent work focused on dictating accurately in

EC 309184

BEST COPY AVAILABLE

conditions similar to the classroom. This primarily involved dictating material from audio and videotapes of classroom lectures and related material. In real-time, as the video (or audio) tape is played, the captionist shadows the material. That is, she/he repeats as closely as possible what the tape says. This part of the work involved considerable learning about what is necessary to remember the spoken material, dictate it into the recognition system and simultaneously listen to new material. The final part of the adaptation work focused upon using the system in the classroom. In the work in the classroom, we used the Stenomask that silences the dictation so that it does not distract students and teachers. By the end of the project, we had used the ASR system in nine college courses.

Findings.

1. Procedures and interfaces. In the first part of this work, we learned about how the microphone used with the system affects accuracy of speech recognition. For example, the Parrott 10 produces a consistent speech signal with a better signals to noise ratio, compared to other microphones that we used at the beginning of the project. In addition we found that variations in background noise in the room in which the individual is dictating can affect accuracy. We then learned to achieve as good accuracy with the Stenomask as with standard speech recognition microphones. We have achieved excellent speech recognition results with the Stenomask, and it functions well in various settings, including those with considerable background noise.

We found that the type of material dictated affects accuracy. Speech recognition works better with standard newspaper or lecture type material, as opposed to literary prose or poetry. Since, in general, the dictation that occurs in the classroom will be this type of standard dialogue, midway through the project we began to work only with these types of materials and are using such material in all training procedures that are being developed.

We learned that in order to achieve high accuracy with the system as soon as possible, the following are important: (a) Immediate, phrase by phase correction of dictated material; (b) repeated training of speech recognition system with same materials until high accuracy is achieved;

(c) work with printed materials, as opposed to audio or videotapes, until high accuracy is achieved for this type of material; and (d) use of appropriate speaking rate and pronunciation during dictation

2. Shadowing.. In the projects initial work at shadowing, we achieved about 95 percent accuracy. Following is an example of the dictated and *corrected* lines for a passage that was done in the second year of the project (First line, regular font—what was dictated; second line, *italicized* is corrected).

> Product to require special handling up often go from the wholesaler to a retailer to
> *Products that require special handling often go from the wholesaler to a retailer to*
>
> the consumer n other places wholesalers are used to help distribute products.
> *the consumer. In other places wholesalers are used to help distribute products.*
>
> This is common for some products. Some products demand a certain type of
> *This is common for some products. Some products demand a certain type of*
>
> distribution system other products can beat the delivered to consumers by a number
> *distribution system, other products can be delivered to consumers by a number*
>
> of different channels.
> *of different channels.*

As can be seen, the text produced with speech recognition contains some errors, but it is generally comprehendible. This segment was "typical" of the performance with speech recognition during shadowing in earlier periods of the project. Some segments had fewer errors and others had more errors.

Dictating taped material in real-time into a Stenomask with high accuracy has proven to be more challenging than originally anticipated. First there is the task of remembering the discourse while simultaneously dictating. Second, the words must be spoken distinctly. In listening to the lecture in real-time, there is a tendency for the captionist to dictate rapidly in order to keep up; but too rapid dictation causes deterioration in pronunciation, especially the endings of words. Third, the captionist must correctly dictate the punctuation (e.g. "period," etc.), and must convert the discourse being heard into appropriate sentences and paragraphs. We have learned that the captionist can more easily perform these tasks when the lecture material is presented on videotape, instead of audiotape. A reason that videotapes are easier is that they provide visual cues, such as

facial expressions, etc. For example, when the lecturer looks at his/her notes, this action may suggest that it is appropriate to insert a new paragraph in the dictation. This work with video and audiotapes is preliminary to the work that will be done live in the classroom.

We have also developed guidelines regarding the extent of correction of mistranslated words it is desirable for the captionist to perform in real-time, and how close to verbatim the captionist may come while shadowing. These guidelines are connected to the speaking rate of the lecturer. (a) When the speaking rate is relatively slow, the dictation will be verbatim, or close to verbatim, and the captionist will correct mistranslations in real-time. (b) When the speaking rate is fast, the dictation of the captionist will include more summarization of the text, and relatively few mistranslations will be corrected in real-time. In addition, the captionist may want to inform the student regarding the extent that the dictated text is a summarization. This may help the student with understanding the relation between the dictated text provided by ASR and the content of the actual lecture.

Subsequent work involved shadowing in the classroom. This is quite challenging, but it can be performed successfully. We can produce now text from ASR that is 97 percent accurate when shadowing in the classroom.

3. Stenomask. The stenomask contains a microphone in a cup that fits over the captionist's mouth, effectively silencing the dictation, so that it does not distract other students in class. We experimented with two versions of the stenomask. For the larger version, the cup containing the microphone covers the nose, as well as the mouth. We achieved accuracy with this type of stenomask that is about one percentage point lower than that achieved with a microphone designed specifically for speech recognition. For example, for one lecture, we achieved 94% accuracy with the stenomask and 96% accuracy with the Parrott Q-10 microphone for speech recognition. We also experimented with a second, smaller version of the stenomask that was found not to work successfully.

After we started doing shadowing in the classroom, we found that the noise produced in shadowing in the classroom may occasionally be distracting to the other students in the class. That is, some of the captionist's voice "leaks" from the Stenomask into the classroom. We have continued to experiment with ways to reduce this leakage of sound.

5. Vocabulary builder and "train-word" dialogue. We worked with the vocabulary builder that adds new words to the system's dictionary and adjusts the system's language model to more closely fit the material being dictated. For example, if the lecture contains specific labels such as "Nisei," and proper names such as "McCoy" and "DeWitt," entry of these words into the system by typing them increases accuracy in subsequent dictation. Entry of text of the lecture, or notes about the lecture, also appear to increase accuracy. Operating the "train word" dialogue, in which the system is familiarized with pronunciation of certain words by repeatedly saying these words to the system until it recognizes them, also appears to increase accuracy, but does not appear to improve it noticeably more than simply typing in the words, and requires more time. We have continued to experiment to determine what is a reasonable amount of information pertinent to the lecture, such as specialized terms that may be used during the lecture, that may be entered prior to the lecture that enhance accuracy when the lecture is actually dictated.

6. Correcting mistranslated words in real-time. If the speech recognition system mistranslates a word, this mistranslation may mislead the deaf student relying on the display for understanding. For example, in the dictation of one lecture, the word, "loads," was mistranslated as "those" in the sentence; "The new factories produced *loads* of pollution." It seems desirable for the captionist to send a signal that indicates a word has been mistranslated, so that the student is aware of it. It also appears desirable to correct the mistranslation as soon as possible. We are continuing to experiment with ways that the captionist may signal that a mistranslation has occurred and to correct to the mistranslated word when there is a pause in the lecture, or another opportunity to correct the word.

7. Training manual. We drafted a preliminary version of a training manual. This 39 page first draft of the manual had 11 sections, including using the Dragon system, improving accuracy, customizing the system to different settings, and preparing C-Print notes. This preliminary manual has provided a basis for subsequent work for a manual that provides instructions on specific steps that one must take in using ASR for shadowing.

8. Lag-time. Because of the concern regarding delay, or lag time, we have compared the lag time for this system with that for a keyboard-based computerized word-abbreviation system that was previously developed by the NTID research and development group and that is widely used in the U.S. as a support service. We measured the time between the instructor's initiation of speaking a set of sentences and the time that the last word for the set appeared on the screen. We did this for sets one, two, and three sentences in length. Measured in this way, lag time for speech recognition is half that for the typing-based system. This suggests that the issue here may be the tendency of the speech recognition system to "chunk" the display of words (i.e. no words displayed, then several displayed at once, rather than one by one or phrase by phrase).

Because of the concern with lag-time, in the fall of 2000, the project purchased a newer, much faster computer, that was a MacIntosh, as opposed to a PC computer and switched from using the NaturallySpeaking to the IBM ViaVoice ASR software. The rate of errors remained about the same, but the lag-time was reduced markedly. In particular, the ViaVoice software displays the text it produces in a steady "phrase by phrase" manner, as opposed to NaturallySpeaking which would often chunk text in two-sentence segments, which meant there would be considerable wait time before the segments would appear.

Evaluation of the Effectiveness of the ASR System

The evaluation had two components. The most extensive part was an experiment with 10 student participants. The experiment and results are presented below. The second part was interviews with three students about their experiences in using the system in the classroom. This

second part is briefly discussed after the experiment. The reasons for conducting the experiment with only 10 students and interviewing only 3 students are presented in the discussion section.

Experiment

This pilot experiment compared the extent that students comprehend and remember information from a short lecture when it is presented with an interpreter in contrast to when it is presented with the text display from ASR. The study also examined students' study strategies in this context. Previous studies that have compared deaf students' comprehension and retention of interpreted and printed information have included real-time displays of text, a hard-copy text, or captioning on television (Gates, 1971; Norwood, 1976; Stinson & MacLeod, 1980; Stinson, Meath-Lang, & MacLeod, 1982; Stinson, et al., 2000,). However, we have not found any studies of students' performance under a real-time speech-to-text system that used ASR.

Purpose The experiment had three goals: (a) To determine whether students retained more information from watching the ASR display and from reviewing a hard copy of the ASR text or from watching an interpreter and reviewing the notes of a notetaker. (b) To determine whether students recalled more information when they are given strategies for reviewing the ASR text or the notes than when they are not given these materials. (c) To obtain students perceptions regarding the quality of information that they receive with ASR.

Design All students viewed one lecture with an interpreter and one with a real-time ASR display, with lecture topic and type of presentation (interpreter/notes vs.ASR) counterbalanced. The format for viewing and studying the lectures was varied for two groups, one with three students, and one with seven students.. The first group was tested immediately after having an opportunity to conduct a review of either a copy of written notes provided by a notetaker or a hard copy of the ASR text. The second group received a sheet that suggested some strategies that students could use in reviewing the notes or ASR text. These students then reviewed the notes or ASR text, and completed the tests after the review activity.

Method

Participants. Participants were 10 college students who were deaf or hard of hearing who attended the National Technical Institute for the Deaf (NTID). Although data were not collected on the degree of hearing impairment and reading proficiency, students in a similar experiment that we conducted at NTID had approximately a 10th grade level on a test of reading comprehension and a pure-tone threshold in the better ear of approximately 95 db. The students in the present study appeared to have similar characteristics.

Materials. The experiment used two lectures of approximately 1,600 words that required about 17 minutes to deliver based on segments of an introductory sociology course that was taught at the Rochester Institute of Technology. One of the lectures was an introduction to sociological concepts and the other was a discussion of social stratification. These lectures were approximately equal in length, vocabulary, number of details, and interest. The "level" of language that was used was typical of that in classes in which the ASR would be used as a support service.

Each of the lectures was videotaped in color with the same female lecturer speaking at a rate of 100 words per minute. For each lecture, a first videotape was made of the speaker alone. Next, at a separate time, a second videotape was made of an interpreter who was a member of the professional interpreting staff at NTID. While the original lecture was being played back, the interpreter used lip movements, signs, and fingerspelling to represent a transliteration of the spoken message, except for omission of a few inflections and function words (Siple, 1995). A third videotape was also made of the real-time display produced with ASR, with operator listening to the original lecture and dictating into the computer. The video camera recorded the real-time text displayed on a computer monitor. This ASR display was produced with one of the early computers used in the project with the NaturallySpeaking software. This meant that there was considerable lag time and that text often appeared in two-sentence "chunks," rather than being steadily displayed in a phrase, by phrase manner.

To produce the materials for review, a hard-copy printout of the material that the ASR operator or "captionist" produced in real time was made. To produce notes similar to those provided by notetakers as a

support service, two trained notetakers independently produced handwritten notes, in accordance with the guidelines of Osgurthorpe (1980). These two notetakers then met to agree on a single set of notes for the study that contained all the critical concepts that contained all the critical concepts for use in the study.

The strategy tip sheet that was used with the strategy group gave suggestions for (a) identifying important information in the notes or text (e.g. "Look for statements that summarize ideas for part of the lecture"); (b) remembering information (e.g. "Skim a section first and then go back and mark the important ideas"); and (c) reviewing and self-checking (e.g. "Review you notes; pay special attention to your underlining and other marks").

Measures. For each lecture, students were given 20 sentence-completion items pertaining to the lecture. For example, one stem item was, "Sociology is the study of human behavior in _____." Students also completed a 15 item multiple-choice test in which each multiple-choice item had four alternatives, such as the following: "Sociologists _____: (a) describe how society works," (b) "judge whether behavior is good or bad," (c) "a and b," or (d) "study individual human behavior." (Correct answer is a). Reliability coefficients (alpha) in a previous experiment with these same materials were .86 (Lecture 1) and .88 (Lecture 2) for the sentence-completion tests and .70 (Lecture 1) and .79 (Lecture 2) for the multiple-choice tests. Prior to taking the tests, the amount of time that students took in studying the notes was recorded.

Students also received a questionnaire with 8 items about their perceptions of the ASR system. These questions asked about the lag time between when the lecturer spoke and when words appeared on the ASR display, the amount of the lecture that they understood, reactions to the errors produced by the ASR, and the extent that students could use context to help deal with occasional errors. The response alternatives for these questions are discussed in the presentation of the results for these questions.

Procedure. All students viewed one of the lecture presentations on two 21-inch television monitors. One monitor showed a playback of the lecturer. The second monitor simultaneous played back in real time either the ASR text or the interpreter. Students were randomly assigned to one of two view/study formats: (a) Review group. For the interpreted presentation, the students reviewed a copy of

the previously composed handwritten notes for 20 minutes after viewing the videotape. For the ASR condition they reviewed a hard copy of the ASR text for 20 minutes after viewing the videotape with the ASR display. For each presentation after finishing the review, students answered sentence-completion and multiple-choice tests. (b) Strategy group. Immediately after viewing the presentation, students in the strategy review group received a sheet that suggested 12 strategies that students could use in reviewing the notes or ASR text. After reading the tip sheet, these students reviewed for 20 minutes the notetaker's notes or the ASR hard copy that corresponded to the presentation mode. They then answered the two tests. For the session with the ASR presentation, students completed the questionnaire about perceptions of the ASR following completion of the objective tests.

Results

Because data were collected for only 10 students, means are reported, but not tests of statistical significance. The first set of analyses focused on performance on the recall and recognition tests. For the comparison of performance for the interpreted and ASR presentations, students performed somewhat better for the interpreted presentation (recognition $\underline{M}$ = .72, recall $\underline{M}$ = .71) than for the ASR presentation (recognition $\underline{M}$ = .65; recall $\underline{M}$ = .59). In the comparison of performance between the group that received notes (in addition to the ASR or interpreted presentation) and the one that received study tips (in addition to the notes and ASR or interpreted presentation), students in the two groups performed at about the same level (notes group, $\underline{M}$ = .66; study tips group, $\underline{M}$ = .69).

The next set of analyses examined the amount of time spent studying the notes provided by ASR or the notetaker before taking the tests. Students in the study tips group (M = 9.50 minutes) spent somewhat more time studying than those in the notes group (M = 7.43 minutes).

The final set of analyses focused on responses to the 8 items in the questionnaire about perceptions of the ASR system. For the item about lag time, students gave a mean rating of 7.25 on a scale that went from 10 ("bothered me very much") to 5 ("some") to 9 ("not at all"). For the two questions about how much they could understand from the interpreted and ASR presentations, students indicated they could understand slightly more from the interpreted presentation. The mean rating was 79.5% for the interpreted

presentation on a scale that went from 0 to 100% understanding and the mean was 72% for the ASR presentation. An additional question asked how much did errors produced by the ASR "bother" comprehension. The mean rating was 5.35 on a scale that went from 10 ("bother me a lot") to 5 ("bother me a little") to 0 ("don't really bother me"). In addition, 6 of the 10 students stated that the errors affected their understanding of the presentation and 3 said that the errors did not bother them. For the question on how much did other words in the sentence help them to figure out the sentence meaning when there was an error, students gave a mean rating of 5.7 on a scale that went from 10 ("very much") to 5 ("some") to 0 ("not at all"). Two questions were open-ended ones that asked for comments. Students indicated that ASR had good potential as a support service. The major problem with ASR was the lag time, and the occasional mistranslation of words by the ASR system was an additional problem that affected comprehension.

Use of ASR in Class and Interviews with Students.

In addition to conducting the experiment, the project used ASR with approximately 30 deaf/hard of hearing students. After class, students regularly received edited transcripts, or "notes" produced with ASR, in which the errors have been corrected. These were distributed by paper, or posted on the web. Anecdotal comments from teachers and students regarding the notes were positive. The teacher used them as a basis for teaching another section of the class. Approximately 10 of these 30 students viewed the real-time display after class. All these students viewed the NaturallySpeaking, ASR text display that had greater lag time, as opposed to the display produced with ViaVoice. We interviewed three of these students. Comments of so few students, at most, provide tentative suggestions. In general, students' comments suggested that the system functions similarly to the keyboard-based word-abbreviation system that the project previously developed. Student feedback yielded two concerns. First, ASR produces too many errors, even with 97% accuracy. Second, there is too long a delay between the time that the teacher says a word during the lecture and the time that word appears on the laptop screen for viewing by the students

Discussion

This section first discusses the work in adapting ASR for use in the classroom. It then interprets the results of the evaluation studies. For both sections, we discuss some of the work that has been done

after the work under Spencer funding was completed. We have learned from the work supported by Spencer that using ASR in the classroom is a formidable task. As is elaborated below, the work has made clear a number of difficulties in using ASR, but it has also pointed to ASR's promise. Because of this promise, we plan to continue the work with ASR, to explore its connections to certain aspects of student learning, in conjunction with using other software, to implement nation-wide training with ASR, and to collect more evaluative data. This work will take at least several years.

Adaptation of ASR for the Classroom

It took much longer than anticipated, but the project has made ASR function successfully in the classroom. Considerable work was done in the project to move from simply using ASR for dictation of printed material to using it in a manner that will function effectively in the classroom. This work was slow, step-by-step, with much trial and error. If we had gone into the classroom without these preparations, it is quite likely that we would not be successful. In doing this work we have found that our previous experience with the keyboard-based word-abbreviation system has been helpful and we used exclusively in the dictation work captionists who had learned the keyboard-based system and had used it in classrooms.

Effective functioning in the classroom requires accomplishment of the following activities: transporting the computers to class, setting up the equipment before class, providing wireless communication with another computer so that the student can view the display, making corrections of the errors in the transcript after class, and distributing the corrected transcript to students as notes. These activities have been shown to be reasonable tasks and indicate that ASR can be an effective support service for deaf/ hard of hearing students.

Students" responses indicate that ASR produces a text display similar to that produced with the typing-based system. From technical and educational perspectives, these have been major goals of the project. This outcome supports our belief that speech recognition can function successfully in the classroom and that it can be an important tool in the education of deaf/hard-of-hearing students.

As the evaluation data discussed below indicate, however, the ASR system has its limitations. From the student perspective, the primary concern is now the errors that ASR produces in the text. Development of better strategies for creating voice files prior to using the system in class, correction of errors in real-time during the class, and more effective use of typing in conjunction with the use of the ASR are three strategies that the project is currently using to reduce errors and improve the comprehensibility of the display.

Work in 2001-2002 Project staff have continued to work with the ASR system in the past year, since June, 2001, when all Spencer funds were exhausted. A major improvement in the system was an additional upgrade in computers, which further reduced lag time for the display of the text. Another improvement has been to use a new mask that more effectively reduces noise produced by the dictation. In April, 2002, our group demonstrated the ASR system at a conference of the Post Secondary Educational Program Network for Deaf and Hard of Hearing Students. The demonstration included use of the ASR system as the speech-to-text support service at eight breakout sessions in the conference. As a result of this demonstration several educational programs expressed interest in trying the system. A workshop on using ASR as a support was held in May 2002 and a few programs have began trying the system. More programs will try the system in the 2002-2003 academic year. Our group at NTID is keeping in contact with these programs to learn about their strategies in using the ASR system in order to further improve it.

Summary. There is still considerable work to do, but we feel that the approach to using ASR that we have developed, in which the Spencer Foundation provided the key support, is well on its way to being widely used by educational programs.

Evaluation Results

Interpretation of results from the experiment. All results are quite tentative because of the small sample size and the lack of tests of statistical significance. The results tentatively suggest that students did better on the recall and recognition tests with the interpreted presentation than with the ASR one. A difference of 10%, which was the difference between the means, could well be

statistically significant if the sample were larger. There are several qualifications, however. First, mean scores might well change if data were collected for 40 or more participants. Second, reading proficiency data should be examined, and the recall and recognition data should be related to the reading scores. Students with higher reading proficiency might do better with the ASR presentation. Third, the experiment was conducted with the NaturallySpeaking ASR display, which chunked information, often in two-sentence units, that may have been difficult to follow. Performance might be better with an ASR display that did not chunk the text in this manner.

The study tips group did not appear to remember any more information than the group that did not receive study tips. One dimension that was not examined, and that should be examined in the future, is the actual strategies that students used. Time spent studying, however, did increase for the study tips group relative to the group that did not receive tips.

Questionnaire responses. The responses to the questionnaire about perceptions of the ASR system suggested that the lag-time, the need to wait for these two-sentence chunks to be displayed as text, interfered with, that is "bothered," comprehension. This finding contributed to the change to the ViaVoice software that was discussed previously.

The mean comprehension rating for the ASR presentation, although lower than that for the interpreted presentation, was only about 7 percentage points lower. This difference is small, and even if we had collected data for 40 students, the difference is not likely to have been statistically significant. The degree of comprehension for the ASR display suggests that the quality of information in the text display is probably similar to that for the keyboard-based system, except for the chunking of text. This issue has been addressed by changing to the ViaVoice ASR software and using faster computers.

Data also suggested that the errors produced by the ASR were bothersome, but not as bothersome as the lag-time. Furthermore, students could use the context of the sentence to help figure out a sentence's meaning when there were errors, but only to a limited extent.

Interview data The interview data for the three students were consistent with the questionnaire data. These data, too, pointed to the lag-time and ASR word mistranslation errors as being important considerations for students.

Limitations of Studies and Explanation of Small Number of Participants.

The major limitation, of course, was the limited number of participants in the experimental and interview studies. These limitations were due to problems in working with the ASR and computer communication technology and to the need to change the technology midway in the investigations. We are disappointed with inability to collect more data during the project.

Experiment. For the experiment, we stopped data collection when it became apparent that the lag-time was interfering with comprehension and the project changed from NaturallySpeaking to ViaVoice. In hindsight, we should have made new videotapes for the experiment as soon as we acquired the new computer and ViaVoice software. However, we did not do this, and collected the data for the 10 students using the older computer and NaturallySpeakingAfter we had done the analysis with the 10 students, we did make new videotapes for the experiment using the new computers and ViaVoice software to produce the ASR text display. In the past year we have collected data for 21 students. Next year we plan to collect data for an additional 35 for a total experiment N of 56.

Interviews. For the interviews a major problem occurred when we switched from using a PC to using a MacIntosh computer, which at the time was the fastest computer that would work with ViaVoice. However, with this switch, the wireless communication between the computer that was producing the ASR and the computer that displayed the text for the student would no longer work. Thus, for the classes in which we used the MacIntosh computer, we were only able to distribute notes after class that were based on the edited text that was produced with ASR during class. As noted, these text notes were much appreciated by students and faculty, but were not really innovative. These notes were very similar to the text-notes produced by the keyboard system that we have distributed and research previously (Elliot, Stinson, Everhart, McKee, & Francis,

2000). Because we did not think that we would acquire much new information from conducting interviews with students about these notes, we did not conduct them.

Summary

The fact that student perceptions as indicated by responses in the interview and experiment questionnaire were in many respects similar to perceptions obtained previously in studies with the keyboard-based system is encouraging (Elliot et al, 2001). We believe that we can improve the ASR system in further work, and, in turn, increase the favorableness of students' perceptions. The data from the experiment and the interview, however, also pointed to the problem of lag-time, which has now been reduced, and to the problem of word mistranslation errors by the ASR system. This area needs the most work in the future, even though the accuracy level now is typically 97%.

References

Elliot, L., Stinson, M., McKee, B., Everhart, V., & Francis, P. (2001). College students' perceptions of the C-Print speech-to-text transcription system. Journal of Deaf Studies and Deaf Education, 6, 285-298.

Gates, R. (1971). The reception of verbal information by hearing-impaired students through television medium- a comparison of speechreading, manual communication, and reading. Proceedings of th 45th Meeting of the Convention of American Instructors of the Deaf, 513-522.

Norwood, M. (1976). Comparison of an interpretd and captioned newscast among hearing-impaired high school graduates and hearing-impaired college graduates. Unpublished doctoral dissertation, University of Maryland.

Osguthorpe, R.T. (1980). The tutor/notetaker. Washington, DC: Alexander Graham Bell Association for the Deaf.

Siple, L. A. (1998). The use of addition is sign language transliteration. In A. Weisel (Ed.) Issues unresolved: New perspectives on language and deaf education. Washington, D.C.: Gallaudet University Press.

Stinson, M.S.& MacLeod, J. (1980, April). Recall of thematically relevant material by deaf students as a function of interpreted versus printed presentation. Paper presented at the meeting of the American Educational Research Association, Boston, MA.

Stinson, M.S., Meath-Lang, B., & MacLeod, J. (1982). Recall of different segments of an interpreted lecture by deaf students. American Annals of the Deaf, 126, 819-825.

Stinson, M.S., Kelly, R., Elliot, L., Colwell, J., Liu, Y., & Stinson, S. (2000, April). C-Print, interpreting, and notes, and memory of lectures. Paper presented at the Annual Meeting of the American Educational Research Associaiton, New Orleans, LA.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").